

# 평면을 이용한 이미지 시퀀스에서의 3D 그래픽 정합에 대한 연구

김주완<sup>o</sup> 장병태  
한국전자통신연구원 공간정보기술센터  
{juwan<sup>o</sup>, jbt}@etri.re.kr

## A Study on 3D Graphics Registration of Image Sequences using Planar Surface

Juwan Kim<sup>o</sup> ByungTae Jang  
Electronics and Telecommunications Research Institute, Spatial Information Technology Center

### 요 약

본 논문은 캘리브레이션 정보를 모르는 카메라로부터 얻은 시퀀스 영상에서 공간상에서 평면인 물체의 영상 정보를 이용하여 카메라 내부 및 외부 파라미터를 추정하고, 이를 이용하여 가상의 3D 그래픽을 시퀀스 영상에 정합하는 방법을 제안한다. 제안된 방법은 기존의 방법에 비해 손쉽게 이미지에 가상의 3D 그래픽 오브젝트를 정합할 수 있으며, 눈에 보이는 정합오차를 최소화하며 DirectX와 같은 3D 그래픽 툴과 쉽게 연동이 되는 장점이 있다. 본 연구는 비디오와 같은 영상에 3D 영상을 합성하는 대화형 비디오 콘텐츠 개발에 활용될 수 있을 것으로 기대된다.

### 1. 서 론

최근들어 실사와 그래픽을 혼용하는 증강현실(Augmented Reality) 콘텐츠 개발에 관련된 연구가 많이 진행되고 있다. 실사와 그래픽 영상을 정합하기 위해서는 실사에 대한 정확하고 신뢰성 있는 빠른 트래킹 기술이 요구되며, 자기트래커, 초음파트래커와 같은 다양한 종류의 트래킹 장비를 이용하여 구현되고 있다. 이러한 방법들은 증강현실 콘텐츠 개발에 별도의 트래킹 장비가 요구되는 단점이 있어, 최근들어 영상 자체에서 트래킹 정보를 추출하는 연구가 많이 진행되고 있다.[1,2,3,5]

컴퓨터 비전 기술의 발전으로 인해 캘리브레이션 정보를 모르는 이미지 시퀀스에서 카메라 모션을 추출하는 것이 가능하나, 이들 작업은 많은 시간이 소요되는 단점이 있으며, 첫프레임을 기준으로 상대적인 카메라 트래킹 정보만을 추출하여 가상의 3D 그래픽 오브젝트를 사용자가 수작업으로 배치해야 되는 문제가 있다.

본 논문에서는 이미지 시퀀스에 있는 평면 정보를 이용하여 상대적으로 사용이 간편한 이미지 시퀀스와 3D 그래픽과의 정합 방법을 제안한다. 기존 방법[5]에서는 카메라 트래킹 정보만을 추출하는 작업에 초점이 맞춰져 있어, DirectX와 같은 툴을 이용하여 3D 그래픽 오브젝트를 정합하기에는 적합하지 않은 구조를 갖는 단점이 있으나, 본 논문에서는 이미지 시퀀스에서 카메라 트래킹 정보를 추출한 다음, DirectX와 같은 그래픽 툴과의 정합을 위해 추가적인 오차 보정을 수행하여 보다 유연한 영상 정합을 이룰 수 있는 장점이 있다.

본 논문의 구성은 다음과 같다. 1장은 서론이며, 2장에서는 본 논문에서 제안하는 방법에 대한 세부적인 절차를 기술하며, 3장에서는 실험 결과를 논하며, 4장에서 결론 및 향후 연구 과제를 논한다.

### 2. 본 론

#### 2.1 개요

두 평면의 대응하는 모든 점들은 호모그래피(Homography) 관계로 표현할 수 있다. 즉, 카메라를 움직이면서 같은 평면을 찍은 두개의 영상에 존재하는 평면에 있는 대응점  $x$ 와  $x'$  사이에는 식 1과 같은 관계가 있다. 이들 관계를 나타내는 변환  $H$ 를 호모그래피라고 한다.

$$x' = Hx \quad H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \quad (\text{식 1})$$

실세계에 존재하는 관심 있는 평면이  $Z=0$  이라면, 이 평면상의 점  $X$ 는  $(X \ Y \ 0 \ 1)$ 로 표시할 수 있으며,  $i$  번째 카메라에 의해 투영된 점  $x'$ 는  $(x \ y)$ 로 표현할 수 있다. 카메라 투영 행렬  $P_i$ 가 3차원 공간상의 점  $X$ 를 이미지상의 점  $x'$ 로 투영하는 행렬이라고 하면, 실세계 평면에 있는 점  $X$ 와 카메라 투영 행렬과의 관계는 식 2와 같다.

$$x' = P_i X = K[R \ | \ t](X \ Y \ 0 \ 1) = K[r_1' \ r_2' \ r_3' \ t'](X \ Y \ 0 \ 1)^T \quad (\text{식 2}) \\ = K[r_1' \ r_2' \ t'](X \ Y \ 1)^T = H_w(X \ Y \ 1)^T$$

여기서  $P$ 는  $3 \times 4$ 의 카메라 투영 행렬이며,  $K$ 는  $3 \times 3$ 의 카메라 내부 파라미터 행렬이며,  $[r_1 \ r_2 \ t]$ 는 카메라 외부 파라미터 중 회전 행렬의 첫째와 둘째 칼럼과 이동 벡터에 해당한다. 이 식에서 보는 바와 같이 호모그래피  $H$ 는  $H = K[r_1' \ r_2' \ t']$ 로 표현할 수 있다. 카메라 외부 파라미터에서 회전 행렬은 직교 행렬이므로, 회전행렬의 세째 칼럼은  $r_3 = r_1 \times r_2$ 로 구할

수 있다.

따라서 카메라 내부 파라미터  $K$ 를 알고 있다면, 매 프레임 별로 카메라 외부 파라미터를 구할 수 있다

캘리브레이션 정보를 모르는 이미지 시퀀스에서 평면을 이용한 3D 그래픽을 정합하는 본 논문의 전체적인 흐름도는 다음 그림 1과 같다.

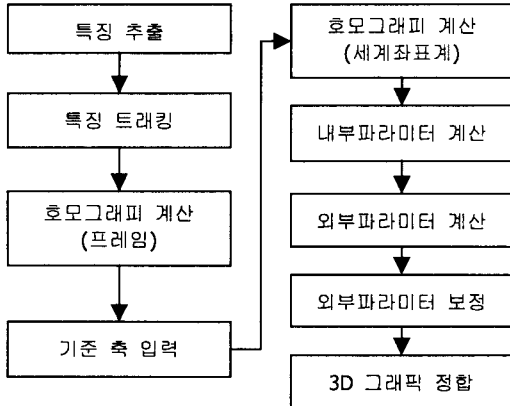


그림 1. 전체적인 흐름도

2.2 특징점 추출 및 트래킹

인접 프레임간의 호모그래피를 구하기 위해서는 프레임간의 평면에 존재하는 대응점의 집합을 찾아야 된다. 본 논문에서는 영상의 그래디언트(Gradient) 값에 의한 공분산 행렬(covariance matrix)  $C$ 를 사용하여 특징점들을 추출하였다. 먼저 영상을 가우시안 필터링하여 그래디언트 이미지  $\hat{I}$ 를 구하고,  $x, y$  방향으로 편미분  $\hat{I}_x, \hat{I}_y$ 를 계산하여, 식 3과 같은 공분산 행렬  $C$ 의 두 고유치(eigenvalue)  $\lambda_1, \lambda_2$ 가 모두 큰 값일 때 특징점으로 추출하였다.

$$C = \begin{bmatrix} \langle \hat{I}_x^2 \rangle & \langle \hat{I}_x \hat{I}_y \rangle \\ \langle \hat{I}_x \hat{I}_y \rangle & \langle \hat{I}_y^2 \rangle \end{bmatrix} \quad (식. 3)$$

$$e = \iint [I(x+d, t+\tau) - I(x, t)]^2 w(x) dx \quad (식. 4)$$

선택된 특징점에 대해 다음 프레임 ( $t + \tau$ )에서 식 4를 최소로 하는 대응점을 구하는 방법으로 특징점을 트래킹하였다. 트래킹 과정에서 손실된 특징에 대해서는 새로운 특징점을 추가하여 항상 일정 수 이상의 특징점들이 트래킹되도록 하였다.

2.3 호모그래피 계산

인접 프레임  $i$ 와  $i+1$ 간에 대응되는 특징점 집합(S)를 이용하여 outlier는 제거하고 남은 inlier 만을 이용하여 프레임간의 호모그래피( $H_i^{i+1}$ )를 다음과 같은 절차에 의해 계산하였다.

단계 1: 대응되는 특징점 집합 S에서 DLT(Direct Linear Transformation)에 의해 호모그래피( $H_i^{i+1}$ ) 계산

단계 2: S의 각 점에 구해진 호모그래피( $H_i^{i+1}$ )를 적용하여 거리 문턱값(Distance Threshold)  $t$  보다 큰 점들은 outlier로

간주하여 S에서 삭제

단계 3: 단계 2 과정에서 삭제된 outlier가 없으면, 구해진 호모그래피를 해당프레임의 호모그래피로 선택

단계 4: 반복 횟수가 N보다 작으면 outlier가 제거된 대응점 집합 S를 이용하여 단계1부터 다시 수행하고, 그렇지 않으면 종료

2.4 카메라 내부 파라미터의 결정

본 논문에서 카메라 내부 파라미터는 서로 직각인 두쌍의 평행선을 이용하여 근사값으로 구하였다. 즉, 카메라의 skew는 0, aspect ratio는 1, 카메라의 principle point는 (0,0)으로 가정하였고, 카메라의 초점거리(focal length) 만을 구하였다. 평면 내에 사각형이 있으면 사각형으로부터 서로 직각인 두 쌍의 평행한 라인을 얻을 수 있다. 평행한 라인에 의해 생기는 두개의 소실점(vanishing point)을  $v = (x_v, y_v, 1)$ ,  $w = (x_w, y_w, 1)$  이라고 하면,  $K^{-1}v$ ,  $K^{-1}w$ 은 직각이므로,  $v^T K^{-1} K^{-1} w = 0$ 의 관계가 성립한다. 이 식으로부터 식 5가 유도되어 카메라의 초점거리  $f$ 를 구할 수 있다.[6]

$$x_v x_w + y_v y_w + f^2 = 0 \quad (식. 5)$$

2.5 카메라 외부 파라미터 및 보정

3D 그래픽 영상을 정합하기 위한 세계좌표계와 이미지 프레임간의 관계를 내부 파라미터의 결정에 사용된 사각형의 좌표값을 이용하여 세계좌표계에서 이미지 좌표계로의 호모그래피( $H_w^0$ )를 DLT를 이용하여 계산하였다. 세계좌표계에서  $i$ 번째 카메라로의 호모그래피는 식 6과 같이 구할 수 있다.

$$H_w^i = H_w^0 H_0^1 H_1^2 \dots H_{i-1}^i \quad (식. 6)$$

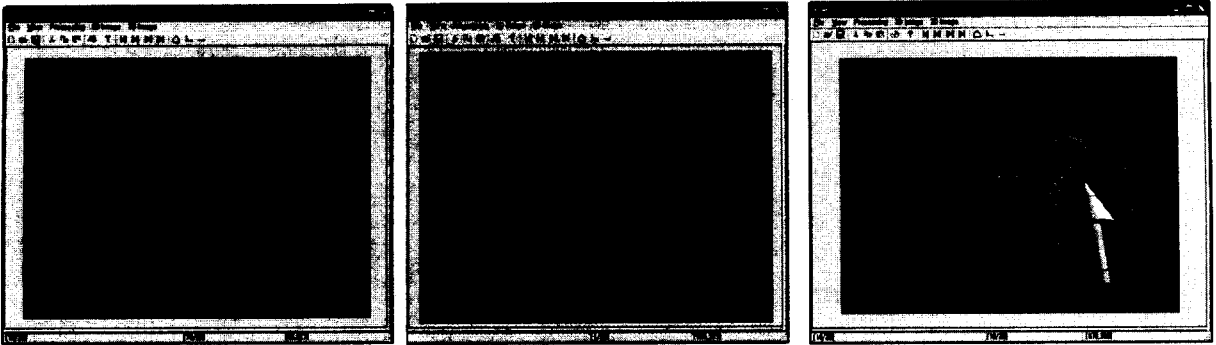
2.4절에서 구해진 카메라 내부 파라미터와 프레임 별로 구한 호모그래피를 이용하여  $K^{-1}H_w^i$ 에서  $[r1 \ r2 \ t]$ 를 구한 다음, 회전행렬이 직교 행렬이라는 성질을 이용하여  $[r1 \ r2 \ (r1 \times r2), t]$ 를 구할 수 있다.

그러나, 위에서 구해진 회전 행렬을 DirectX에서 바로 적용할 때 오차가 있다. 특히 실제 구해진 회전 행렬  $R = R^T R = R R^T = I$ 인 직교행렬의 성질을 만족하지 못하여, 3D 그래픽 오브젝트를 렌더링하면, 3D 그래픽 오브젝트가 찌그러져 보이는 현상이 발생한다. 따라서 본 논문에서는 회전 행렬  $R$ 을 SVD(Singular Value Decomposition)를 이용하여 직교 행렬의 성질을 만족하는 회전 행렬  $R'$ 로 식 7을 이용하여 보정하였다.[4]

$$UDV^T = SVD(R) \quad (식. 7)$$

$$R' = U * I(3,3) * V^T$$

이와 같이 구한 회전 행렬  $R'$ 은 직교 행렬의 성질을 만족하지만, 보정되는 과정에서 원래의 축과 어긋나게 되어 3D 그래픽 오브젝트와 정렬 오차를 발생시킨다. 따라서, 평면에 있는 사각형의 좌표를 이용하여 카메라 외부 파라미터 중에 이동 벡터를 조정하여 세계좌표계상의 좌표축과 이미지상의 사각형의 좌표값과의 차이가 최소가 되도록 최소자승법을



(가) 특징 추출 및 트래킹

(나) 오차 보정 전

(다) 오차 보정 후

그림 2. 단계별 결과 영상

이용하여 보정하였다.

### 3. 실험 결과

2장에서 기술한 방법에 따라 구현하고, 바닥에 신문이 있는 동영상에 이용하여 실험을 실시하였다. 그림 2의 (가)에서는 평면이 있는 이미지 시퀀스에서 특징 추출 및 트래킹을 수행한 결과 영상이다. 프레임간에 대응되는 특징점을 아웃라이어와 인라이어로 분리한 다음, 인라이어만을 이용하여 프레임별로 호모그래피를 계산하였다. 신문의 테두리 정보를 이용하여 카메라 내부 파라미터를 계산하였으며, 신문의 테두리는 3D 그래픽을 정합하기 위한 기준축으로 사용하였다. 그림 2의 (나)는 구한 외부 파라미터에 대해 회전 행렬을 근사한 직교 행렬로 보정한 다음 기준 축과의 정합된 결과이며, 그림 2의 (다)는 이동 벡터를 조정하여 오차를 보정한 후, 3D 그래픽을 정합한 결과 영상이다. 그림에서 보는 바와 같이 기준 축이 신문의 테두리 영상과 거의 일치하는 모습을 보여 주고 있다. 이미지상의 그리드는 3차원 공간상에서 Z=0인 평면을 0.5 길이의 격자를 투영 행렬에 의해 투사한 결과이다.

보정하였으며, 그림 2의 (다)에서 보는 바와 같이 X축과 Y축의 길이가 같게 된다.

### 4. 결론

본 논문에서는 캘리브레이션 정보를 모르는 이미지 시퀀스를 입력받아 이미지 내에 존재하는 평면 정보를 이용하여 가상의 3D 오브젝트를 정합하는 방법을 제안하였다. 그리고, 본 논문에서 기술된 방법에 따라 MS DirectX를 이용하여 기존의 방식에 비해 보다 정확하게 이미지 시퀀스에 3D 그래픽을 정합한 결과를 보였다.

제안된 방법에서는 캘리브레이션 정보를 두쌍의 직교하는 평행선을 이용하여 근사화된 카메라 캘리브레이션 방법을 적용하였기 때문에 3D 그래픽이 기준 좌표축에서 멀어지면 멀어질수록 정렬 오차가 커질 수 있으며, 평면이 없는 이미지에서는 사용 할 수 없는 단점이 있다. 향후에는 보다 정확한 캘리브레이션 방법의 연구와 평면이 없는 일반적인 이미지 시퀀스에도 쉽게 적용할 수 있는 연구가 필요하다.

표 1. 기준축을 역투영(Reprojection)한 결과

구분	기준 좌표 (픽셀)	거리오차
이미지상의 기준축	(241,390) (534, 358) (228,134)	-
보정전	(226.3,405.5) (516.2,376.6) (214.1,148.9)	33.4
보정후	(240.6,390.8) (534.6,357.2) (186.3,178.5)	0.89

표 1은 기준축의 이미지 좌표값과 3D 공간상의 좌표축을 첫프레임에 대해 투영행렬 보정전,후의 행렬로 재투사한 결과값을 보여주고 있다. 기준축은 원점(0,0,0), X축(1,0,0), Y축(0,1,0)의 값이며, 원점과 X축에 대한 보정전의 거리오차는 33픽셀이었으나, 보정후의 거리 오차는 0.9픽셀로 정합 오차가 줄었다는 것을 보여준다. Y축의 길이는 기준축의 형태가 정사각형이었다는 것을 가정하여  $K^{-1}H_v$ 의 첫 두열의 frobenius norm의 비로 나누어 길이를

### 5. 참고문헌

[1] R.Azuma, J.W.Lee, B.Jiang, J.Park, S.You and U.Neumann, "Tracking in unprepared environments for augmented reality systems," IEEE Trans. On Computer Graphics, 1999.  
 [2] P.A. Bearsley, P.H.S.Torr and A. Zisseman, "3D model acquisition from extended image sequences," In Proc. ECCV, pp. 683-695, 1996.  
 [3] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2000.  
 [4] Gene H. Golub, MATRIX Computations, The Johns Hopkins University Press, 1996.  
 [5] Gilles Simon, Andrew W. Fitzgibbon and Andrew Zisserman, "Markerless Tracking using Planar Structures in the Scene," Proc. International Symp. On Augmented Reality, Oct 2000.  
 [6] B. Caprile and V. Torre, "Using vanishing points for camera calibration," IJCV, pp.127-140, 1990.